DOCUMENT RESUME

ED 116 666 IR 002 910

AUTHOR Crumpton, John
TITLE Data Processing for Educators; a Vocabulary Self-Study and an Application Review. Monograph No. 1.
INSTITUTION Durham Technical Inst., N.C.
PUB DATE 75
NOTE 49p.
AVAILABLE FROM Durham Technical Institute, P.O. Box 11307 East Durham Station, Durham, North Carolina 27703

EDRS PRICE MF-$0.76 HC-$1.95 Plus Postage
DESCRIPTORS Autoinstructional Aids; Computers; *Computer Science Education; *Data Processing; Educational Technology; Models; Programing; Programing Languages; *Textbooks

ABSTRACT

This brief text serves as an introduction to the vocabulary, concepts and application of data processing. It is intended to provide the professional educator with a working vocabulary in the areas of: (1) unit record data processing; (2) data processing systems; (3) computer-centered data processing; (4) programing; (5) approaches to data processing; and (6) decision-making models. The text is designed to be self instructional, and it contains periodic questions to test the learner's comprehension. (EMH)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED116466

DATA PROCESSING FOR EDUCATORS

A VOCABULARY SELF-STUDY AND AN APPLICATION REVIEW

by JOHN CRUMPTON, Ed D.

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

"PERMISSION TO REPRODUCE THIS COPYRIGHTED MATERIAL HAS BEEN GRANTED BY
Randell S. Cain, Durham Technical Inst. & Author
TO ERIC AND ORGANIZATIONS OPERATING UNDER AGREEMENTS WITH THE NATIONAL INSTITUTE OF EDUCATION. FURTHER REPRODUCTION OUTSIDE THE ERIC SYSTEM REQUIRES PERMISSION OF THE COPYRIGHT OWNER."

IR 002 910

Copyright © by John Crumpton, 1975

TABLE OF CONTENTS

	Page
INTRODUCTION	iii
CHAPTER 1 - UNIT RECORD CENTERED DATA PROCESSING	1
CHAPTER 2 - SYSTEMS	8
CHAPTER 3 - COMPUTER CENTERED DATA PROCESSING	11
CHAPTER 4 - THE STORED PROGRAM AND PROGRAMMING LANGUAGES	19
CHAPTER 5 - APPROACHES TO DATA PROCESSING	24
CHAPTER 6 - D. P. APPLICATIONS FOR EDUCATION	30
CHAPTER 7 - THE DECISION-MAKING MODEL	43
BIBLIOGRAPHY	45

INTRODUCTION

The purposes of this book are to strip the cloak of mystery from data processing for the professional educator and to provide him with information and tools to help make valid decisions about data processing as it relates to his special environment. The method is simple; to give him an opportunity to relate data processing concepts with data processing vocabulary, and then to help him remove data processing from the abstract and place it in context with education.

To aid the educator in retaining the factual and conceptual information contained in this document, periodic "benchmarks" are taken in the form of "QUESTIONS TO BE ANSWERED." They serve the dual purpose of improving retention and directing the user to areas needing restudy.

The materials are intended to serve either as a "mini-text" or to be self-teaching for an audience ranging from juniors and seniors in schools or colleges of education through educators with many years of professional experience. The result may be overexposure for many who use the materials. That is as should be expected; however, there is nothing contained herein that is beyond the limits of understanding of the members of the defined audience.

Chapter 1 - Unit Record Centered Data Processing

As in any special environment, data processing requires the potential user to have a fundamental understanding of data processing, itself, before there is a possibility of its being the object of intelligent decision-making. In the interest of economizing the reader's time brevity will be the watchword.

Data processing can be classified into two broad categories, Unit Record Centered and Computer Centered. It is important to note that the distinction between the two categories is not sharply separated in certain instances because of the tendency of smaller computers to function using unit record techniques. Unit Record Centered data processing will be examined first, since it is historically an earlier data processing phenomenon than Computer Centered data processing and because of its being conceptually less difficult.

The Unit Record is the now-familiar punched card. It is a standard sized (Gold Standard dollar bill sized), standard data volume (80 characters of information), document for the recording and processing of data. The standardization of physical size and physical record length is valuable in and of itself, but the really magnificent value is that The Unit Record has its data recorded in a manner that allows a machine to easily and accurately read that data.

The machines that typically utilize The Unit Record are called Electronic Accounting Machines or Unit Record Machines and are, with certain exceptions, nothing more than large elec-

tronic calculators with the ability to be programmed (carry out a well-defined set of instructions, repetitiously) through physically changing circuitry pathways with a "patch panel" or "plug board." The calculating ability is usually limited to addition and subtraction of dollars and cents along with the ability to carry forward totals and to crossfoot on a ledger or journal operation. Special devices are occasionally attached that allow multiplication and division of dollars and cents (at a pace that seems painfully slow when compared to the modern computer).

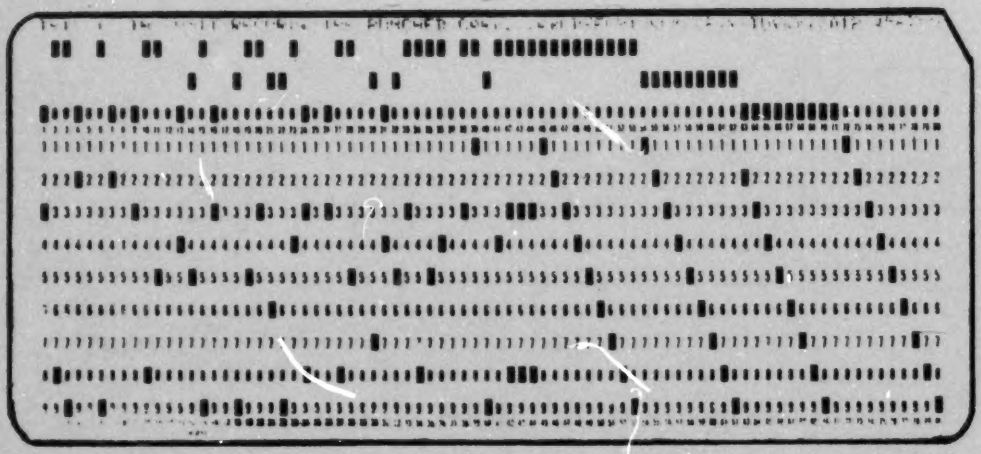
Most Unit Record Machines are equipped with some ability to print information contained in or derived from the punched card unit record. The typical print mechanism is divided into print positions or print columns. The number of positions limits the useable width of the printed page; for example, a machine with only 50 print positions would be limited to a total print width of about five inches, but a machine with 132 print positions could print a line as much as 13 inches wide! One print position is the location in which an alphabetic character, a number, or a special symbol may be printed.

Since most Unit Record Machines print an entire line at a time with type bars or type drums (composed of print wheels) located in each print position, the printing capability (variety of characters in each position, print width, and print speed) is limited by the type font on each type carrying device (bar or wheel), the number of print positions that are provided, and the ability of the machine to achieve frequent positioning

of the print carrying devices in the proper locations to print a line of information. Print speeds vary up to about 250 lines per minute with conventional Unit Record equipment.

Other machines that operate on The Unit Record are sorters, which sequence the cards in some orderly fashion (typically a numeric or alphabetic sequence), collators, which match and/or merge two or more groups of punched cards (for example, inserting name cards in front of a collection of cards containing payroll information about that person, but which is identified only by an employee number), interpreters, which read the data in the card and print that information directly on the card itself, reproducers, which read an existing card and punch all or part of its information in another card or cards, and of course the keypunch, which is the device used by an operator to convert information from the usual manually prepared document to The Unit Record. Each of these machines is programmable.

At this point it is desireable to discuss the basics of using The Unit Record. First of all, The Unit Record is a single physical record containing 80 characters or columns of information. Often one finds the need for a record whose length is greater than 80 characters (as in the case of the payroll record mentioned earlier), so two or more unit records (punched cards) are combined to form a logical record of more than 80 characters length. Of course there would be little advantage to using unit record data processing (or any other form of data processing, for that matter) if there was only



one logical record so there are usually many logical records combined to form a file. Files are usually compartments of like data all of which relate to common unit record sub-systems or systems.

A typical unit record system might be a personnel system, composed of sub-systems such as a payroll sub-system and a personnel information sub-system. These two sub-systems may share a name and address file, composed of logical records each containing the name and address of an individual employee.

A contrasting approach from the viewpoint of the record is a situation in which the 80 character physical record contains more than one logical record. Looking more closely at our friend, the name and address file, the name and the street number may both be on the same card along with an employee number to associate it with other records and files. The name

may be considered one logical record, the street number may be considered one logical record, and the employee number may be considered one logical record.

The concepts of physical and logical may be applied to files as well as to records. A physical file may consist of a metal cabinet full of punched cards. The metal cabinet may contain a logical file for payroll, a logical file for an inventory control function, and a miscellaneous collection of other logical files. It is equally possible, however, for a physical file to contain only a part of a logical file: imagining the entire Internal Revenue Service income tax file in punched cards conjures visions of thousands of physical files to represent a single logical file.

Briefly defining the concepts just discussed:

1. The Unit Record is usually considered to be the traditional 80 column (80 character), large dollar-bill sized, document with which we are all familiar.

   NOTE: Recently available products have introduced a new version of The Unit Record that contains 96 characters of information in a data processing card about the size of a credit card.

2. A physical record is the actual medium on which data is recorded; i.e., the punched card.
3. A logical record is a collection of logically related information (without regard for characteris-

tics of the medium).

4. Unit Record Machines include large electronic calculators with limited arithmetic and often with line printing capabilities. Other machines perform transactions on (or against) The Unit Record.
5. A logical file is a collection of related logical records.
6. A physical file is the physical medium or media that contain a given logical file, a part of a logical file, or a collection of several logical files.
7. A system may be considered to be a collection of interrelated, interacting, logical files.
8. A sub-system is a system that is totally contained within another system.

It is important to note that The Unit Record approach to data processing is remarkably unsuited to a highly "scientific" environment, but is quite at home in the "business" environment.

QUESTIONS TO BE ANSWERED:

This is your big chance to prove how well you have read the materials thus far presented.

1. What is The Unit Record? What do you believe its importance to be?
2. What do you believe to be the differences between logical and physical as they relate to The Unit Record?
3. Briefly describe the functions of some of the machines

that are used to process The Unit Record.

4. What is a system? Can you describe some broader implications of the systems concept for data processing than are presented in the preceding materials?

When you complete your answers to the questions, carefully evaluate your answers to questions 1, 2, and 3 by comparing them against the content of the section. Question 4 will be discussed in Chapter 2.

Chapter 2 - Systems

The concept of a system was briefly examined and partially developed in Chapter 1. It is so important to data processing, however, that it deserves more attention. We are, therefore, going to spend this chapter examining it from several points of view.

A system was defined as a collection of interrelated, interacting, logical files. It is, indeed, that, but it is also much more.

The place to begin looking at the concept of system is with the job or application for which data processing is being used. In general a system is a major operational part of the business or institution being examined.

A familiar system (and one near and dear to our hearts) is the payroll system. It is composed of an attendance recording (or keeping) application, a vital statistics application (job title, wage, deductions, etc.), a warrant (or check) writing application, a warrant distribution application (payday), a banking application, a check reconciliation application (balancing the bank book), and a complaints from the employees application (error detection/correction). All of these applications interact with each other to form a functional system that solves (or attempts to solve) the payroll problem. They also interact with applications from other systems (the banking application with the receivables-payables system).

The full view of the system includes the people who make it work, the people subject to its services, the business flow,

and the data processing activities that support it. The system constitutes everything that goes into making it work.

Systems are result oriented. They pay the employee; they control the inventory; they order the merchandise; they collect the bills; they pay the vendors; they place the resources (personnel, equipment, money); they solve the problem.

In data processing in the Unit Record Centered environment or in the Computer Centered environment, the system is key. Success or failure depends on how well the data processing systems designer can fit the data processing system to the existing business or educational system and how well he can evolve the existing system into a condition where it comfortably associates with his data processing system.

The systems designer (systems analyst) must understand the "industry" in which he is working as well as the limits and capabilities of the data processing system at his command. He must merge his industry and data processing into a harmoniously functioning whole.

There is an alternative view of system (or perhaps an additional view) as the concept applies to data processing. In this very closed environment the term may be applied to the machinery and its very specialized support, such as programming (which will be discussed later).

Data processing machinery is often referred to as hardware, and in the data processing vernacular it is not unusual to hear the expression, "hardware system," or a modular set of a particular data processing equipment vendor's hardware and special-

ized support as something like "System/950."

Within the data processing world, system also refers to all the things to be done and to be used in completing a particular application or set of applications. The entire spectrum of data processing within a specific organization is often referred to as a "Total System."

QUESTIONS TO BE ANSWERED:

1. What is your idea of the concept of system?
2. How might the data processing person understand the concept of a system? How does it differ from the opinions you expressed in answering question number 1, above?
3. What is an application? How is it like a system? How is it different from a system? How does it relate to a system?
4. How might a Computer Centered System differ from a Unit Record Centered System?

When you complete your answers to the questions, review Chapter 2, comparing your answers to questions 1,2, and 3 to the text. Keep your response to question 4 in the forefront of your mind while reading the following chapters.

Chapter 3 - Computer Centered Data Processing

Computer Centered data processing invariably uses The Unit Record to some extent. The more completely Computer Centered a specific data processing installation becomes, however, the less The Unit Record or punched card is used.

The most apparent differences between unit record centered data processing and computer centered data processing are in the realms of calcualting capability, files and file media, devices for capturing and distributing data, and SPEED. The most emphatic thing is speed. Computers and their associated devices are FAST.

The computer, itself, is our first area of interest. Another name for the computer is central processing unit (or CPU). The CPU controls the flow of data to and from other devices. It stores data temporarily in the task of receiving and sending data. It performs logical and/or mathematical operations against data it recieves.

The word FAST is most emphatically demonstrated in the CPU. Many recent computers can perform a simple caluclation (add, subtract, multiply, or divide with numbers of GREAT magnitude) in only a few billionths of a second. Some can receive data, process against it, and send it on to another device in less than a millionth of a second. With that kind of speed, fast seems inadequate, so perhaps FAST is, indeed, the correct term.

The CPU's arithmetic ability, generically speaking, is really limited to addition, but by using special additive techniques it can simulate subtraction, multiplication, and division. In fact

it can find solutions to complex problems in calculus, it can invert matrices, and literally solve the most time consuming and tediously involved of mathematical problems, without fatigue and in very short amounts of time.

The physical files used in computer centered data processing also tend to be quick. The Unit Record is still in use, but it is typically read or punched in greater volumes in less time by the computer (reading over 1,000 cards per minute in some cases).

The physical file associated with computer centered data processing with which we all tend to be most familiar is the magnetic tape and the magnetic tape drive. The magnetic tape reel comes in many sizes. All of them from the "Mini-reel" to the "Maxi-reel" (the conventional 12 inch reel) are capable of containing the data from literally thousands of punched cards.

Data on tape can be very dense in terms of how much can be located on a given length of tape. Data density on tape varies from about 200 characters (2-1/2 cards) per inch to 1,600 characters (20 cards) per inch on commonly used equipment.

The data density is a function of (is dictated by) the quality of the tape, itself, and by the electrical and mechanical limitations of the tape drive. The finest tape will acquire a data density of only 200 characters per inch when used on a tape drive that is limited to that density. By the same token a tape recorded at 200 characters per inch may be read only through using a tape drive with a 200 character per inch rating. A poorer quality tape (for example rated or tested at 500 characters per inch) will probably result in errors being recorded when used on

a tape drive designed for greater density (for example 1,600 characters per inch).

The ability of a tape drive to transfer data from tape to the CPU tends to parallel its ability to handle tapes of greater or lesser data density. That is, a 200 character per inch rated tape drive usually (but not always) transfers data to the CPU more slowly than an 800 character per inch rated tape drive.

The transfer rate of a tape drive is a function of data density and of the tape transport speed (how fast the tape drive can move the tape past the device that reads the information stored on the tape). If a high tape transport speed is combined with high data density, a high transfer rate results. For example, at the same tape transport speed a 1,600 character per inch rated tape drive will transfer data to the CPU at twice the rate of another drive with the same tape transport speed but only an 800 character per inch rating. Speeds of a quarter-million characters per second transfer rate are not too uncommon, although about thirty-thousand characters per second would probably be the most common transfer rate for tape drives.

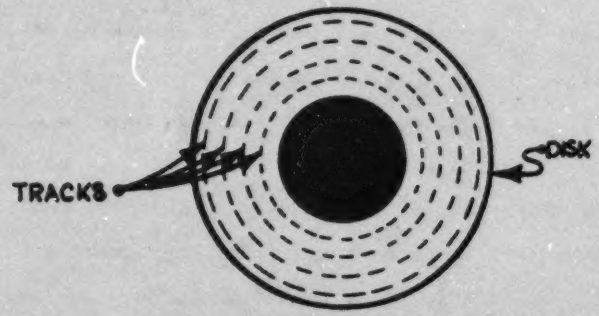
The methods in which data is stored in The Unit Record is through punched holes in the card. The method in which data is stored on magnetic tape is through a magnetic pattern induced in a ferrous (iron) oxide coating on one surface of the tape. The punched hole in the unit record is permanent, so that a change in the data in a given record is accomplished by discarding the old card and replacing it with a new card with the new data. The magnetic record can, however, be easily erased, allowing new data

to be written in the place of old, which of course offers some obvious potential for economy.

There is a disadvantage to magnetic tape as a storage medium, even though it is generally superior to punched cards. Magnetic tape is a long strip with records stored, one after the other (serially), from one end to the other. Many applications (for example, payroll record searches to respond to telephone inquiries) require the examination of individual records in sequences that are not at all like the sequences in which the records are likely to be placed in the storage medium (random or direct access). To randomly access a single record on a magnetic tape file may take from one-thousandth of a second to as much as several munutes, with the avarage time being in the vicinity of four to five minutes, because the tape must be read from the first record until the desired record is found, for each search. It is ovbious that under these circumstances if there are more than 15 random searches (accesses) per hour, the demand for service will exceed the hardware system's capability (and your very expensive computer will do little more than to conduct long, tedious, searches).

To meet the requirement for direct access a variety of direct access storage devices (DASD) have been developed by the many data processing equipment vendors. A majority of these devices use a stack of ferrous oxide coated disks (that look much like a carefully spaced stack of phonograph records) as the place to record data (as the physical file).

Unlike the phonograph record with its spiral gooves containing "data," the data on disks are arranged in concentric circles ranging outward from near the center. These concentric circles are

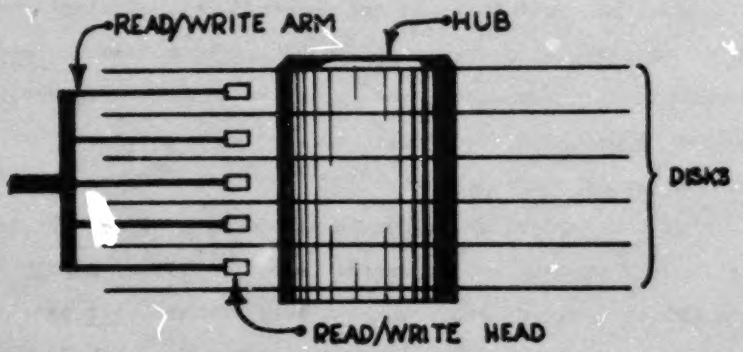


called tracks. Each track, regardless of its diameter, contains the same number of "data positions" as do the other tracks on the same disk. Data is simply more compressed on the more central tracks; that is, it is more dense.

The stacks of disks may have from two disks to one-hundred or more disks. Some stacks are removable modules (much as a reel of tape is removable from a tape drive) and are called disk packs. A disk pack is used on a disk drive. Disk packs are physically interchangeable between the drives for which they are designed. The disk pack may contain a number of logical files or may contain only one or a part of one logical file.

Data on the tracks on the disks are written or read by read/write heads that are similar in purpose to those of the read/write mechanism on a tape drive. There is at least one read/write head for each surface, upper and lower, of each disk (with the occasional exception of the top and the bottom disks where only the inside face has a read/write head). The read/

write heads are mounted on the ends of a device that looks like a very large, heavy, comb, called the read/write arm. The read/write arm positions the read/write heads exactly over the tracks of data, allowing the location of a single item of data in little more than the time required to move the arm into position and the time required for the disk to rotate one-half revolution (one-eighth



of a second is a normal amount of time to find and read from or write to a specific logical record, which results in a capability for as many as 420 inquiries per minute or 25,200 per hour - a considerable improvement over the 15 per hour of the tape mechanism). The normal transfer rates for disk drives is in excess of 150,000 characters per second.

Another high speed, direct access, device containing physical files in magnetic form is the magnetic drum. The magnetic drum is typically several times as fast as the magnetic disk in directly accessing a given logical file. It is also typically several times as expensive as the disk drive, while containing only a fraction of the data that may be contained on disk. Both, however, contain

data in the millions of characters range.

Obviously, with all of these high speed magnetic files, some way of rapidly printing the results of processing data is required. The high speed line printer is the device specially designed for that purpose. Speeds in excess of 1,000 lines per minute of printed copy are common, with a print width as great as 132 characters. That amounts to about thirteen and one-half feet of single spaced copy per minute. With spacing it could easily exceed 50 feet of copy per minute. Even this great speed is often inadequate, so it is not uncommon to find two, three, or even more printers serving a single computer system.

All of the devices described in association with computer centered data processing (other than the CPU) either supply data to the CPU or recieve data from it. They either input data to the CPU or output data from it. The generic term for such devices is Input/Output devices. They are usually referred to as I/O devices or simply as I/O.

The amount of productive data processing resulting from inputing, processing, and outputing is called throughput. High throughput implies a great deal of productive work in very little time. Low throughput implies some sort of "log jamb" in the processing system. The responsible data processing professional is constantly concerned with improving throughput.

QUESTIONS TO BE ANSWERED:

1. Is the Unit Record of use in Computer Centered data pro-

cessing? If so, in what way? If not, why not?

2. What kinds of processes can a CPU perform against data?
3. What is meant by fast in computer centered data processing?
4. Describe the basic characteristics of three different magnetic files.
5. What is the importance of the "stored program" in computer centered data processing?

When you complete your answers to the questions, review Chapter 3, comparing your answers to questions 1,2,3, and 4 with the text. Keep your answer to question 5 in mind while studying Chapter 4.

Chapter 4 - The Stored Program and Programming Languages

The stored program is a phenomenon associated with the digital computer. A program, as was implied when discussing Unit Record Machines in Chapter 1, is a well-defined set of instructions to be carried out by the equipment and the personnel operating the equipment. A sample program for a computer to follow would resemble the very much simplified and literalized set of statements shown below:

001 TURN ON THE CARD READER.
002 READ THE INFORMATION IN THE FIRST LOGICAL RECORD IN THE CARD THAT IS AT THE READ POSITION.
003 TURN ON THE MAGNETIC TAPE DRIVE.
004 READ THE INFORMATION IN THE FIRST LOGICAL RECORD ON THE MAGNETIC TAPE.
005 COMPARE THE TWO RECORDS.
006 IF THE INFORMATION IS THE SAME (MATCHES) EJECT THE CARD INTO THE FIRST STACKER AND GO TO STEP NUMBER 008.
007 IF THE INFORMATION IS DIFFERENT EJECT THE CARD INTO THE SECOND STACKER.
008 IF THERE IS A CARD AT THE READ POSITION IN THE CARD READER GO TO STEP 002.
009 TURN OFF THE CARD READER.
010 TURN OFF THE MAGNETIC TAPE DRIVE.
011 END OF PROGRAM.

This very simple program would compare two records and select those

cards that did not match the magnetic tape record for deposit in a separate output station from those cards that do match the magnetic tape record. Obviously, there is nothing done in this program that couldn't be done by hand in a few seconds for each transaction (using some medium other than magnetic tape for the control file). There would of course be many errors if the job were done by hand, and if there were several thousand transactions, the job would take several hours for a clerk to complete. Data processing equipment could, however, complete the job with complete accuracy within a few minutes time.

The key word at this point is stored when dealing with the stored program. With unit record equipment the programming capability is very limited (although an operation similar to this one could be programmed for unit record systems using cards instead of magnetic tape for the control file) and is accomplished through "rewiring" the machines circuitry with its plug board (patch panel). The computer on the other hand has memory that can be loaded with instructions and data from literally any input device that can be electronically attached to the computer. The instructions (or programs) and information (or data) are electronically stored in the computer's memory, with either data or programming easily and readily changed at the discretion of the programmer or of the system's operator.

With the unit record program the length and involvement were severely limited by the physical capacity of the wiring panels to a dozen or so different instructions and about the same number of steps to a given program. The largest computers have the capa-

bility of responding to over two-hundred different basic instructions and may have primary memory capacity of several million character positions and auxiliary memory (storage facilities like magnetic disk and (magnetic tape) of several billion character positions. The primary memory is usually composed of ferrite cores and is casually referred to as core memory or simply as core. The character positions may be filled with program instructions or with data (or with both if one looks at all character positions available to the computer system).

The stored program adds great versatility to data processing. It allows a large repertoire of instructions as well as programs containing long sequences of instructions. It allows many physical and logical files to interact with the system and allows a tremendous variety of I/O devices to be attached to and controlled by the CPU. The stored program is relatively easy to develop and once developed is inexpensive to retain and maintain and is quite easy to use. The stored program is the key to the remarkable success of computer centered data processing and has contributed significantly to the development of the data processing industry.

The stored program must ultimately be reduced to electrical "statements" that the CPU understands. There is a problem, though, in that people and CPUs do not "speak" the same language. The CPU understands the on-off relationships of Boolean Algebra, and in effect uses algebra as a language (called machine language). Man, as a rule, doesn't understand or care to understand Boolean Algebra, but he does understand English, or Chinese, or French, or German, or Russian, or Japanese, and so on.

In order to overcome the "communications gap" between man and computers, special languages, called programming languages, have been developed. Each language serves a special purpose (for example, FORmula TRANslation language (FORTRAN) would seem to be directed toward the solution of problems associated with mathematics and the sciences, whereas, COmmon Business Oriented Language (COBOL, would seem to relate to business and commerce. Some languages try to satisfy both the scientific and business programming functions and seem audacious enough to claim it in their names; the classic example of such a universal language is Programming Language/1 (PL/1).

The programming language is (to both the CPU and man) a foreign language, so the man must learn the programming language he plans to use as a programmer the there must be some provision for translating the programming language into something the algebra speaking computer can understand. The learning problem for man is typically accomplished in traditional ways. The computer is slow to learn, however, so special translation programs (called assemblers or compilers) are developed to convert the program written by man in a programming language into machine language. The more the programming language looks like man's language, the more likely is its translation program to be called a compiler. The more the programming language looks like machine language, the more likely is its translation program to be called an assembler. Compilers translate languages that resemble man's language (usually English). Assemblers translate languages that are close to machine language. Compiler languages are usually easier for man to learn and to use

than are assembler languages.

QUESTIONS TO BE ANSWERED:

1. What do you understand a stored program to be? How is it similar to and different from the program used with unit record equipment?
2. What are some advantages of the stored program?
3. What is a programming language and what purpose does it serve?
4. What are compilers and assemblers? What advantages are there to man in using a compiler language as opposed to using and assembler language or machine language?
5. What is teleprocessing? How does it increase the benefits of data processing to its users?

When you complete your answers to questions 1,2,3, and 4, review Chapter 4 to compare them with the text. Keep your answers to question 5 in mind while studying chapter 5.

Chapter 5 - Approaches to Data Processing

There are two basic approaches to data processing (when looking at the way in which data are captured and distributed), local and remote. Local data processing means that a source document in some manually prepared form (typewritten, pencilled, quill and inked) is the medium through which the source data or information needed for a particular application arrives at the data processing facility. The data are then, usually, converted to a form that can be read by the data processing equipment through use of a keypunch and key verifying operation, resulting in an accuracy checked stack (deck) of punched cards (some recently developed key entry equipment enters data directly on a magnetic tape instead of punching cards). The deck of cards (or the tape) is then, typically, placed on the appropriate input devices and enter the job stream (a series of applications to be processed in some orderly sequence), is eventually subjected to processing, and a report of some sort is generated (or a scientific problem is solved). Local data processing is more generally referred to as batched job processing or simply batch processing. All unit record data processing is batched job processing.

Batched job processing is also quite common in computer centered data processing. The computer is (as was implied in Chapter 3) much quicker than the I/O devices that may be attached to it; thus, if only one job is running at a time, the CPU spends most of its time waiting for a card to be read, or for a magnetic tape record to be read or written, or for a record to be located and read or written on a magnetic disk file, or for some other I/O

device to do its job. Wait time is usually discontinous, none-the-less, it became obvious that all the waiting was costing money, so a technique called multi-programming was developed.

Multi-programming is conceptually quite simple. All that is done is to take the CPU wait time from one program and run another program in it. If there is still wait time available, a third, or fourth, or fifth program is added until all of the wait time is gone and the CPU is completely utilized. This often means adding more I/O devices, but if you have enough work to justify a second "single-program-at-a-time" (single job stream) computer system, just putting the I/O from the new system on your old computer (if it's a member of the so-called third generation) and using multi-programming may well do the job just as well and save the cost of the second CPU.

Remote data processing means that a device is located at the source to capture data and is directly connected to the computer (without the middle man keying the data into punched cards) by some sort of "solid" electrical connection. "Solid" means something like two strands of insulated copper wire or the use of the facilities of the Communications Common Carriers (the telephone and telegraph companies).

Special hardware devices are usually required in connecting (interfacing) the solid electrical connection to the CPU and to the distant device (terminal). The signals used to represent data along the solid electrical connection (line) are quite unlike the signals used to represent the signal in the CPU and at the terminal. The signals entering the line must be modulated

into a form acceptable to the line, and signals leaving the line must be demodulated into a form acceptable to data processing equipment. The devices that accomplish this task are called modulator/demodulators or MODEMS.

If there is more than one terminal (or line, for that matter) attached to the CPU there must be some sort of facility for receiving data form all of the terminals and separating it, identifying it, and giving the person at the terminal the impression that he has the computer to himself. This task is called multiplexing. There is also a problem associated with the fact that even though the data signals at the terminal and in the CPU are the same, there is generally a distinct difference in the way in which data is expressed within those signals. The devices which perform the multiplexing task and organize the incoming and outgoing data from the CPU are called Transmission Control Units (TCU). When doing remote processing it is usually desireable to simultaneously do batch processing since remote processing usually uses very little CPU time. Multi-programming is of greatest value when the data processing user is mixing local and remote processing.

There are several categories of remote data processing some of which are described below:

1. Remote Job Entry. A completed and tested program and its associated data are placed in the input device of a terminal. The CPU "calls" the terminal and "asks" if it has any messages to send. The terminal replies, "yes," and sends the program and the data. The CPU receives the pro-

gram and data, placing it at the end of its job stream (or input queue). The program and its data wait their turn for processing, are processed, and the results are placed in line (output queue) for being returned to terminals. When the results are due to be transmitted back to the terminal, the CPU "calls" the terminal, "asks" if it is ready to receive its results, the terminal answers "yes" or "no," and if yes sends the results, if no holds it until it can, by re-calling the terminal, get a "yes" answer.

2. Inquiry-update. An operator at a keydriven (typewriter or television-like) terminal receives a request from a person (over the telephone, perhaps, concerning something like the amount of a gas bill). There is a need for an immediate answer (the customer is irate over his gas bill). The operator asks the computer if she may inquire of its files (bids for attention). The CPU responds as promptly as it can (a second or two). The operator then keys as much information as she has about the inquirer (name, address, telephone number, or what ever else is recognized by the computer program). The CPU searches its files and displays all close fits to the entered data (if a name was given, for example, William Woberts, the system may be designed to give the addresses of all the William Woberts and close homonyms in its files along with the account number that corresponds to that name and address). The operator compares the returned information with information known

about or asked of the inquirer ("What is your address, Mr. Woberts?") and then identifies the correct file to the CPU (sends the proper account number). Ultimately, through a series of interactions the desired information is found (the correct billing amount) and the inquirer is satisfied (the bill was in error). In the process the operator may learn that the name was misspelled (Roberts, not Woberts) and enters a change (update) to correct the spelling.

3. Time Sharing. This is usually a problem solving situation. The user in some remote location needs to have the full power of a large computer to solve mathematically complex problems. The user enters, through a keyboard, one line of a program at a time. The CPU reviews each line (statement) and indicates errors in syntax (line-by-line syntax error correction), allowing immediate correction of errors in program statements. Once the program is completely entered, the user enters appropriate data, tells the CPU to process and gets results. The programming languages used by the person at the terminal in time sharing applications are called interactive languages. The significant characteristic of time sharing is that the user demands the attention of the CPU and performs his problem solving task as if he were the only user, even though there may be hundreds of other users operating under the same illusion of total possession of the CPU. Time sharing is timely with respect to the user. A special case within time

sharing is Computer Assisted Instruction (CAI). CAI uses an interactive language for course authors to write interactive student programs. The range of program styles as viewed from the student's perspective varies from the simplicity of Programmed Instruction (PI) to the sophistication of tutorial instruction.

QUESTIONS TO BE ANSWERED:

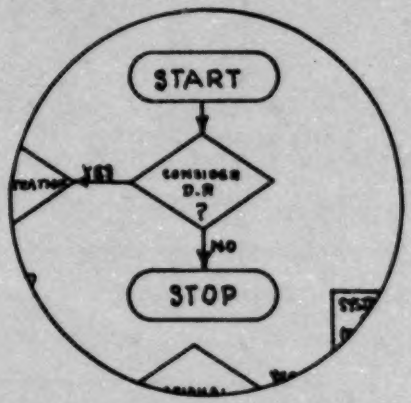
1. Briefly describe local data processing.
2. Briefly describe remote data processing.
3. What are the significant differences between local and remote data processing?
4. Can both local and remote data processing occur on the same computer system at the same time? If so, through what medium? If not, why not?
5. What are the principal applications areas for data processing in the education environment?

When you complete your answers, review answers 1,2,3, and 4 against the text of Chapter 5. Keep your response to answer 5 in mind when studying the next chapter.

Chapter 6 - D.P. Applications for Education

At this point it is assumed that the reader has established for himself a vocabulary that encompasses the content of the first five chapters. That vocabulary and its associated concepts will be extensively utilized in this chapter.

The first decision to be made by the educator is whether there is sufficient justification to seriously consider implementing data processing in his educational institution. Before making that

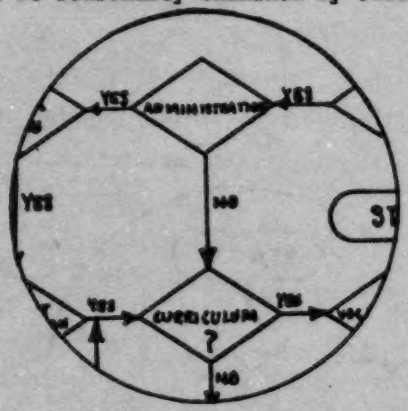


decision, the educator should read Computers in Higher Education, the Report of the President's Science Advisory Committee, The White House, Washington, D.C., February, 1967 (available for 30 cents per copy from: Superintendent of Documents, U.S. Government Printing Office, Washington, D.C. 20402). The document addresses higher education and its data processing needs, but the administrator in the public schools should readily relate the content and recommendations to his situation.

If the decision, once the contributing factors have been

weighed is no, then simply stop and rest in the light of the simple decision. If the answer is yes, then a new question arises; what primary applications areas will be addressed through data processing technology?

There are two basic classifications of data processing applications that can be reasonably examined by educators, those that

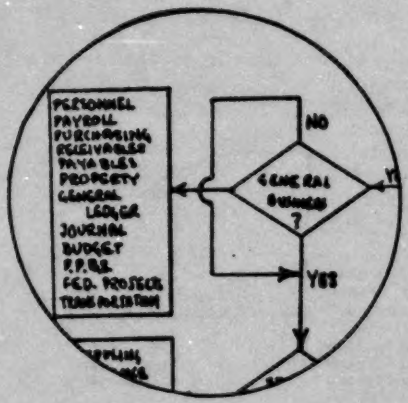


help administer to the institution and those that help teach the students. Those that help administer are, logically, called administrative and those that help teach are called curricular. The decision to do either requires a closer look at both before arriving at any definite conclusions.

Administrative data processing introduces two additional decision areas that deserve consideration. These areas have to do with the kinds of data to be processed and the manner in which the processing of the data will be approached. We are of course referring to applications dealing with the conduct of the business side of education and of applications dealing with the processing of student information; in other words, the General Business

Administration and the Student Administration applications for the institution.

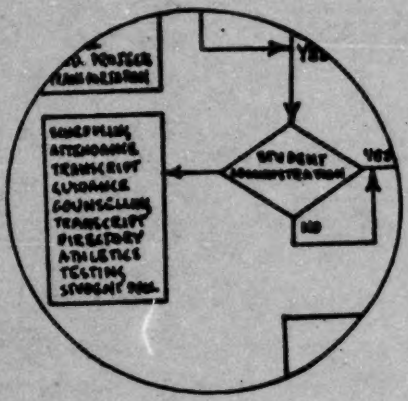
The General Business area incorporates the realm of traditional



data processing, applications dealing with payroll, personnel, recievables-payables, purchasing, property accounting, budget, and on-and-on through the roster of common accounting functions.

There is, typically, very little difficulty in justifying two or three dollars per student month in this area, alone, simply because there are normally personnel and facilities savings that result from implementing data processing that will offset that amount of expense. There are some start-up costs that will be unrecoverable through later cash savings, but once in full operation, the data processing facility for general business applications will usually have little effect on the institution's operating budget, if it is well planned and properly implemented by knowledgeable educational administrators and skilled data processing personnel.

The Student Administration area addresses the problems of student scheduling, attendance accounting, permanent records (and temporary records, too), guidance and counselling, testing, transcripting, student directory, athletic records and accounting, student elections, clubs, and so-on as far as the imagination is willing and able to explore.



Justification for data processing to support these applications is usually a little bit more difficult than for general business applications. The justifications tend to be in the realms of staff-faculty morale, availability of professional personnel to do professional jobs rather than clerical chores, in conservation of physical facilities, in availability of information. A well designed attendance application will almost completely relieve teacher personnel of the attendance accounting function. An effective student scheduling application will conserve as much as one man year per 1,000 students per scheduling effort (each year, semester, quarter, term) of administrator time with fringe benefits such as

accurate class rosters, machine prepared teacher grade books, automatic updating of student records, and considerable savings of teacher clerical time in scheduling. Good design in guidance and counselling applications have in some instances effectively doubled the time available to guidance personnel to perform the guidance and counselling functions with students through virtual elimination of clerical chores. A comprehensive student record system will replace hundreds of file cabinets required for records storage with a few reels of magnetic tape and will allow records access in comparatively short periods of time for either a single active record or a cross section of the institution's entire student population, accessing specific data items or entire records.

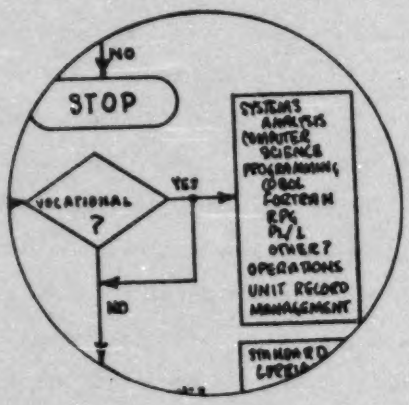
For example in a large public school district a supervisor with exceptional education responsibilities may need to determine the number of students with each I.Q. score below 70 in order to satisfy a reporting requirement for a federally funded project. The manual task, if it were possible, might take weeks or months if he must contact each school within the district, but with a centralized data processing facility containing complete student records, the delay could be as little as a few minutes and could certainly be accomplished overnight.

A well planned, complete, system of data processing applications for student administration will undoubtedly impact the institution's budget. It can, however, be a very inexpensive means of improving the institution's ability to provide quality education to its students.

The curricular applications are more varied than administrative

and are often far more esoteric (and more expensive). There are three curriculum areas requiring consideration, vocational, problem solving, and instruction.

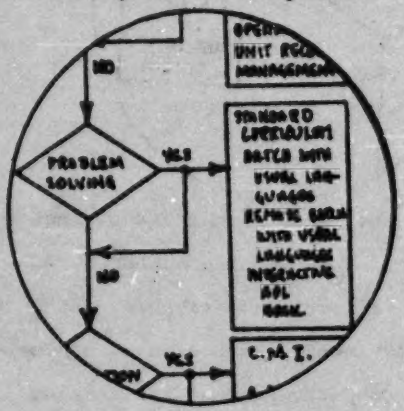
Vocational data processing education tends to quickly bear fruit, although training young people and adults to do data processing requires very capable instructors and usually competes with the sciences in per pupil expense. Curricula that should be examined within the vocational data processing purview include systems analysis (the relating of data processing to an applica-



tions area), computer science (the design of computing equipment and computer programming systems), programming languages (COBOL, FORTRAN, PL/1, etc.), data processing management and administration, unit record systems operation, and computer systems operation. Most vocational training in data processing would be offered to adults. Juniors and seniors in high school are good prospects for programming and operation. Computer science is generally a four year plus graduate degrees college curriculum. A statistic that supports implementation of vocational data pro-

cessing curricula is the more than 100,000 new jobs each year that are being generated by the growth of the data processing industry.

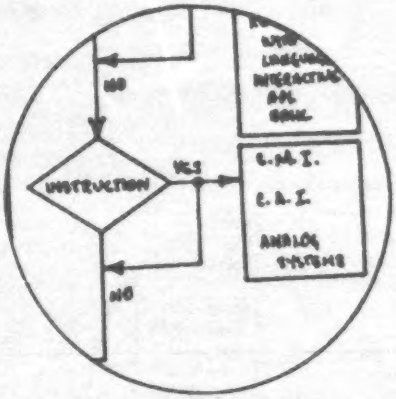
Problem solving is implemented to support the standard curriculum. For example, an accounting class may find COBOL a useful tool in developing journal and ledger accounting techniques. A statistics class could use FORTRAN and a statistical subroutine program to do the arithmetic associated with various statistical techniques, expanding the statistics curriculum from a class in arithmetic into something far more emphatic.



Although the standard curriculum can be supported through local batch processing, an institution with more than usual resources or with an unusually great student requirement for data processing may look to remote batch processing and remote interactive processing for additional computing convenience and power.

Applications that serve to instruct or to direct the instruction of students are generally very costly and are now largely experiments. They can reasonably be expected to remain in the experimental category until well into the 1980s and perhaps be-

yond. The problem of speeding instructional applications does not lie with data processing technology, but with the inability of education to utilize, effectively, the data processing resources that are presently available.

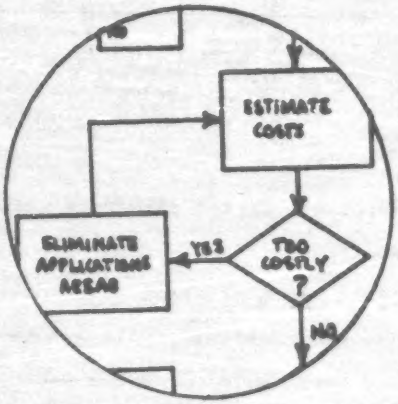


In some special cases, such as the use of analog computer systems in the sciences and the use of computer based simulations to provide students with realistic experiences, instructional data processing is fairly advanced. In the cases of Computer Assisted Instruction (CAI) and Computer Managed Instruction (CMI), there is a great amount of development that must yet take place in the "how to use it" category before the typical educational institution would have reason to seriously consider either, except as interesting curiousities.

As a general consideration in the previous areas, data processing capabilities in the realms of simulation and forecasting should be carefully examined. The ability to simulate student flow through a school building can help in designing more efficient and less costly school buildings. The ability to simulate several

hundred different pay schedules for teachers, arriving at total predictable costs to within a few dollars of what the actual amount will be, can be a significant tool in financial planning. The ability to accurately forecast student population growth by geographic areas would make a far simpler task of physical plant planning, transportation planning, teacher and non-instructional personnel recruitment planning, etc.

Once each of the applications areas that this document describes has been examined and the decisions made as to which seem attractive, it is essential that a cost estimate be acquired. The best way to get a reliable cost estimate is to contact one or two of your local data processing vendors, tell them what you propose

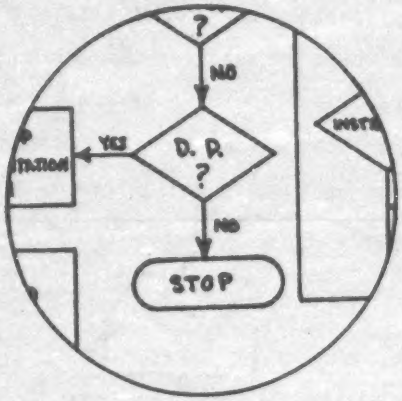


to do, and ask that they suggest machinery to do the job and give you a cost estimate. After receiving the cost estimate, add two times the estimated hardware rental to the hardware rental to allow for personnel, utilities, supplies, and facilities.

RENTAL + 2 x RENTAL = ESTIMATED COST.

It would be simpler to multiply the hardware estimate by three, but the impact is less emphatic than the longer method. If the estimated costs are excessive, eliminate some of the applications that have been planned and seek a new estimated cost.

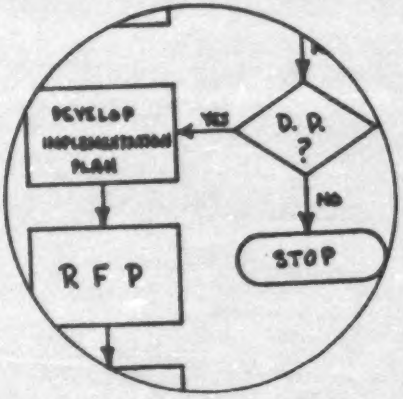
When the cost estimating cycle is complete the administrator should, again, ask himself whether or not data processing is important enough to justify the cost. If the answer is no, then



rest with the easy decision. If the answer is yes, an implementation pla should be developed.

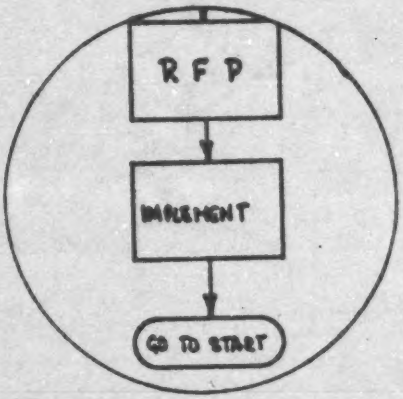
The development of an implementation plan includes deciding on what data processing and clerical people are needed and starting the hiring process, deciding on a review procedure for the establishing of data processing priorities and for resolving conflict relating to data processing, establishing a line to top level administration for data processing management to report through, developing timetables for implementation, and designing an RFP (Request for Proposal) to be submitted to the vendors so that they may bid on equipment along with a valid method of evalu-

ating those bids in the form of proof of performance, benchmarks, and systems service support.



The RFP is an important document. It should be prepared after the data processing manager, his assistant manager for systems analysis, his assistant manager for programming, and his assistant manager for operations have been employed. Its purpose is to tell potiential vendors the things that are to be done by the institution with data processing and advise them of the restraints that are to be placed on bids in terms of costs and of the general characteristics of the hardware (such as specifying disk and tape as physical files). The responses to the RFP should be evaluated in terms of the vendor's ability to do the job that is required. Often the least expensive system will be the most costly because of greater programming costs. It is also, usually, considered a good idea to acquire all of your computer centered equipment from one vendor (to reduce inter-vendor conflict). Once the RFP is evaluated and a contract is negotiated with the vendor(s), the task of implementation must take place.

Implementation means to carry out the implementation plan. The programs must be written and documented, the work flow for data processing must be formalized and evaluated, the physical facility



must be prepared, personnel must be acquired and trained, equipment must be installed, and a conversion from the existing way of doing things to the data processing way of doing things must be accomplished.

QUESTIONS TO BE ANSWERED:

1. What are the two administrative applications areas that are described in Chapter 6?
2. What are the three curriculum applications areas that are described in Chapter 6?
3. How does one go about getting a reliable cost estimate for data processing?
4. What is an implementation plan and what is its value?
5. Mentally move through all the steps described for making

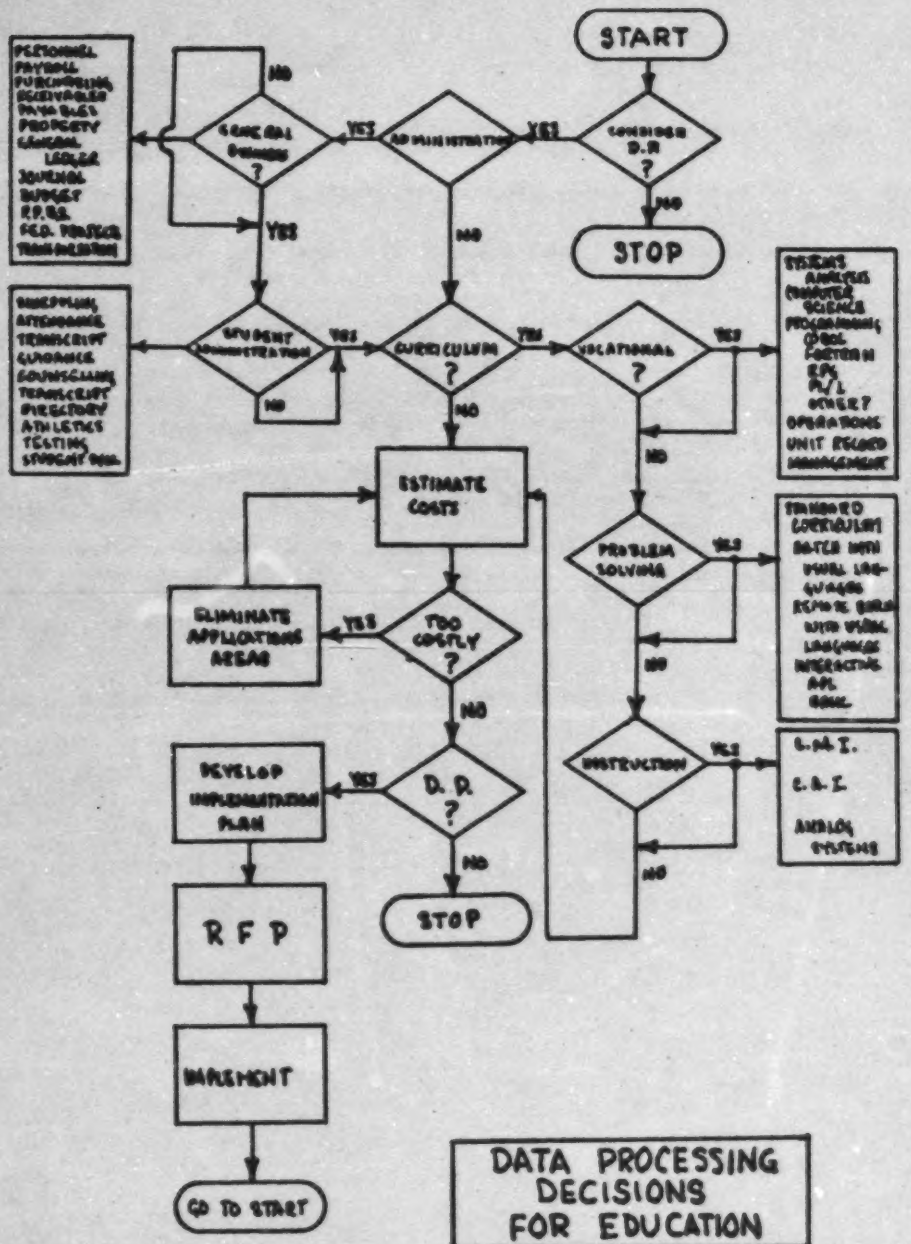
data processing decisions for education. Make a model that relates the many considerations involved.

When you have completed your answers, compare your results for questions 1, 2, 3, and 4 with the text of Chapter 6. Retain your answer to question 5 as you examine the model represented in Chapter 7.

Chapter 7 - The Decision-making Model

Bits and pieces of the model for "Data Processing Decisions for Education" have been presented throughout Chapter 6. It is displayed in its entirety on the following page.

Although there are many more considerations about data processing that a well informed administrator should know, this document serves as an acceptable introduction to the fundamentals. Its readers should be able to carry on an intelligent conversation with data processing professionals and be able to understand the principal decisions that must be addressed when data processing becomes a primary concern.


START
CONSIDER D.P. ?
STOP
ADMINISTRATION
GENERAL BUSINESS ?
PERSONNEL
PAYROLL
PURCHASING
RECEIVABLES
PAYABLES
PROPERTY
GENERAL LEDGER
JOURNAL
BUDGET
R.F.P.S.
FED. PROJECTS
TRANSPORTATION
STUDENT ADMINISTRATION
CURRICULUM ?
VOCATIONAL ?
DISCIPLINE
ATTENDANCE
TRANSCRIPT
GUIDANCE
COUNSELLING
TRANSCRIPT
DIRECTORY
ATHLETICS
TESTING
STUDENT BODY
BUSINESS
ANALYSIS
COMPUTER SCIENCE
PROGRAMMING
COBOL
FORTRAN
RPG
PL/1
OTHER?
OPERATIONS
UNIT RECORD
MANAGEMENT
ESTIMATE COSTS
PROBLEM SOLVING
TRAINED LOOKUPS!
DATA INTO
VISUAL LANGUAGES
REMOTE DATA INTO VISUAL LANGUAGES
INTERACTIVE APL
BASIC
ELIMINATE APPLICATIONS AREAS
TOO COSTLY ?
INSTRUCTION
C.A.I.
C.A.I.
ANALOG SYSTEMS
DEVELOP IMPLEMENTATION PLAN
D.P. ?
STOP
R F P
IMPLEMENT
GO TO START
YES
NO
DATA PROCESSING DECISIONS FOR EDUCATION

BIBLIOGRAPHY

Glans, T. B., Burton Grad, David Holstein, W. E. Meyers, R. N. Schmidt. 1968. Management Systems. Holt Rinehart and Winston, Inc., New York.

Weinberg, G. M. 1966. PL/1 Programming Primer. Mcgraw-Hill Book Company, New York.

Pierce, J. R., et al. 1967. Computers in Higher Education, Report of the President's Advisory Committee. U. S. Government Printing Office, Washington, D. C.

Wilson, J. G., Marthann E. Wilson. 1967. Information, Computers, and System Design. John Wiley & Sons, Inc., New York.

Popell, S. D., et al. 1966. Computer Time-Sharing. Prentice Hall, Inc., Englewood Cliffs, New Jersey.

Meadow, C. T. 1967. The Analysis of Information Systems. John Wiley & Sons, Inc., New York.

Rath, G. J. 1966. Punched Card Data Processing. Science Research Associates, Inc., Chicago, Ill.